

ALLOCATING DATA TRANSMISSION RESOURCES IN PACKET-SWITCHED DATA TRANSMISSION

BACKGROUND OF THE INVENTION

5 [0001] The invention relates to allocating data transmission resources in packet-switched data transmission and especially to optimizing radio interface resources in wireless packet-switched data transmission.

10 [0002] Third-generation mobile systems, called UMTS (Universal Mobile Telecommunication System) and IMT-2000 (International Mobile Telephone System), will provide not only circuit-switched speech services, but also packet-switched services for instance in the manner of the packet radio network GPRS (General Packet Radio Service) designed for the GSM system. Packet-switched data transmission enables the use of different data services by means of a mobile station and, on the other hand, the allocation of the resources of a mobile system, especially the radio interface, for each user as
15 necessary.

[0003] When the user of a terminal in a UMTS system wants to use a packet-switched application, for instance to download a video file from the network to the terminal, the radio resource management system RRM of the UMTS system allocates to a radio bearer an application-based capacity
20 reservation which not only depends on the used application but also on the available radio bearer parameters. In typical one-way data transmission, for instance when downloading a file from the network, a data rate of x bit/s can then be allocated to a terminal in the downlink direction (from the base station to the terminal) and a data rate of 0 bit/s in the uplink direction (from the
25 terminal to the base station). In such an application, the uplink data transmission is typically not needed and consequently, there is no need to allocate resources to it.

[0004] One of the parameters defining the radio bearer is the method used by the terminal for compressing the header fields of data
30 packets. Header compression of data packets being transmitted and decompression of data packets being received is performed on the packet data convergence protocol layer PDCP of the UMTS system. The PDCP layer of a terminal typically supports several header compression methods so as to enable connection set-up with as many network layer protocol types as
35 possible. Some header compression methods may also need a reverse

connection for making different acknowledgments and solving error situations. A certain bandwidth then needs also to be reserved for the reverse connection, but, on the other hand, the compression of the header field decreases the need for a bandwidth for a forward connection.

5 **[0005]** A problem with the above arrangement arises from using an application-based capacity allocation together with a header compression method which requires a bi-directional connection. If the terminal only has header compression methods requiring a bi-directional connection available to it and the terminal sends a capacity allocation request to an application which
10 is typically uni-directional, such as the downloading of a file from the network described above, the radio resources management system RRM only allocates a one-way connection for the radio bearer on the basis of the application. The available compression methods do not then function and the connection cannot be set up without being able to reserve an adequate
15 bandwidth also for the reverse connection using special arrangements comprising new capacity allocation requests. This is not possible in all situations and, in any case, such an arrangement complicates an optimal allocation of radio resources.

BRIEF DESCRIPTION OF THE INVENTION

20 **[0006]** It is thus an object of the invention to develop an improved method and an apparatus implementing the method so as to reduce the drawbacks mentioned above. The object of the invention is achieved by a method and a system which are characterized in what is stated in the independent claims. Preferred embodiments of the invention are set forth in
25 the dependent claims.

[0007] The invention is based on selecting the compression method to be used on the radio link already before setting up a radio bearer, and this information is transmitted to the radio resource management system RRM which takes into consideration in the capacity allocation a possibly selected bi-
30 directional compression algorithm and allocates the necessary capacity for both the downlink and the uplink direction.

[0008] The method and system of the invention provide the advantage that header compression methods requiring a bi-directional connection also function immediately on radio bearers which on account of to
35 the application would only need a one-way radio bearer. A further advantage

is that no extra signaling is needed to set up the reverse connection, because the bi-directional nature is already taken into consideration when the radio bearer is set up. Yet another advantage is that the total bandwidth being allocated to the radio bearer can be optimized when taking into consideration both the bandwidth required by the compression method and the benefit derived from the use of the compression algorithm as a smaller amount of data.

BRIEF DESCRIPTION OF THE FIGURES

[0009] In the following, the invention will be described by means of preferred embodiments, with reference to the attached drawings in which

Figure 1 shows a block diagram of the structure of the UMTS system,

Figures 2a and 2b show protocol stacks used for UMTS control signaling and user data transmission,

Figure 3 shows a block diagram of an functional model of a PDCP layer, and

Figure 4 shows capacity allocation signaling of an embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0010] In the following, the invention will be described using a packet radio service according to the UMTS system as an example. The invention is, however, not restricted to the UMTS system only, but can be applied to any packet-switched data transmission method whose capacity allocation of packet data connections requires that the used header compression methods are taken into consideration.

[0011] The structure of the UMTS mobile telephone system is described in Figure 1. Figure 1 only contains the blocks essential for explaining the invention, but it is obvious to a person skilled in the art that a conventional mobile telephone system also comprises other functions and structures, which need not be described in greater detail herein. The main parts of a mobile telephone system are a core network CN, a UMTS mobile telephone system terrestrial radio access network UTRAN, and a mobile station or user equipment UE. The interface between the CN and UTRAN is referred to as Iu and the air interface between UTRAN and the UE is referred to as Uu.

[0012] UTRAN typically comprises several radio network subsystems RNS, the interface between the RNSs being referred to as Iur (not shown). An RNS comprises a radio network controller RNC and one or more base stations BS, also referred to as nodes B. The interface between the RNC and B is referred to as Iub. The base station BS typically takes care of radio path implementation and the radio network controller RNC manages at least the following: management of radio resources, control of handover between cells, power adjustment, timing and synchronization, paging the subscriber terminal.

[0013] The core network CN is made up of an infrastructure belonging to a mobile telephone system and external to UTRAN. In the core network, a mobile switching center / visitor location register 3G-MSC/VLR is connected to a home location register HLR and preferably also to a service control point SCP of an intelligent network. The home location register HLR and the visitor location register VLR comprise information on mobile subscribers: the home location register HLR comprises information on all subscribers in a mobile network and the services they subscribe to, and the visitor location register VLR comprises information on mobile stations visiting the area of a certain mobile switching center MSC. A connection to a serving node of a packet radio system 3G-SGSN (Serving GPRS Support Node) is formed through an interface Gs' and to a fixed telephone network PSTN/ISDN through a gateway mobile switching center GMSC (not shown). A connection from the serving node 3G-SGSN to external data networks PDN is formed through an interface Gn to a gateway node GGSN (Gateway GPRS Support Node) which has a further connection to the external data networks PDN. The connection from both the mobile switching center 3G-MSC/VLR and the serving node 3G-SGSN to the radio network UTRAN (UMTS Terrestrial Radio Access Network) is set up through the interface Iu. It should be noted that the UMTS system is designed in such a manner that the core network CN can be identical to the core network of a GSM system, for instance, in which case there is no need to rebuild the entire network infrastructure.

[0014] The UMTS system also comprises a packet radio system which is to a large extent implemented according to a GPRS system connected to a GSM network, which explains the references to a GPRS system in the names of the network elements. The UMTS packet radio system can comprise several gateway and serving nodes, and several serving nodes

3G-SGSN are typically connected to one gateway node 3G-GGSN. Both nodes 3G-SGSN and 3G-GGSN function as routers supporting the mobility of a mobile station, which routers control the mobile system and route data packets to mobile stations regardless of their location and the used protocol.

- 5 The serving node 3G-SGSN is in contact with a mobile station MS through the radio network UTRAN. A task of the serving node 3G-SGSN is to detect mobile stations capable of packet radio connections in its service area, to transmit and receive data packets from said mobile stations and to track the location of the mobile stations in its service area. Further, the serving node
- 10 3G-SGSN is in contact with the mobile switching center 3G-MSC and the visitor location register VLR through the signaling interface Gs' and with the home location register HLR through the interface Gr. Records related to packet radio services and comprising subscriber-specific packet data protocol contents are also stored in the home location register HLR.

- 15 **[0015]** The gateway node 3G-GGSN acts as a gateway between the UMTS network packet radio system and the external data network PDN (Packet Data Network). External data networks include the UMTS or GPRS network of a second network operator, the Internet, an X.25 network or a private local area network. The gateway node 3G-GGSN is in contact with said
- 20 data networks through the interface Gi. Data packets being transmitted between the gateway node 3G-GGSN and the serving node 3G-SGSN are always encapsulated according to the gateway tunneling protocol GTP. The gateway node 3G-GGSN also contains PDP (Packet Data Protocol) addresses of the mobile stations and routing information, i.e. 3G-SGSN addresses. The
- 25 routing information is thus used to link the data packets between the external data network and the serving node 3G-SGSN. The network between the gateway node 3G-GGSN and the serving node 3G-SGSN employs an IP protocol, preferably the IPv6 (Internet Protocol, version 6).

- 30 **[0016]** Figures 2a and 2b show UMTS protocol stacks used for control signaling (control plane) and user data transmission (user plane) in a packet radio service of the UMTS system. Figure 2a shows the protocol stack used for control signaling between a mobile station MS and the core network CN. Mobility management MM of the mobile station MS, call control CC and session management SM are signaled on the highest protocol layers between
- 35 the mobile station MS and the core network CN in such a manner that the base stations BS and the radio network controller RNC located in between are

transparent to this signaling. Radio resource management of radio links between mobile stations MS and base stations BS is managed by a radio resource management system RRM which transmits control data from a radio network controller RNC to base stations BS. These functions related to the general management of a mobile system form a group called core network protocols (CN protocols), also known as Non-Access Stratum. Correspondingly, the signaling related to radio network control between a mobile station MS, a base station BS and a radio network controller RNC is done on protocol layers called radio access network protocols (RAN protocols), i.e. Access Stratum. These include transfer protocols on the lowest level, whose control signaling is transferred to the higher levels for further processing. The most essential of the higher Access Stratum layers is the radio resource control protocol RRC which is responsible for establishing, configuring, maintaining and releasing radio links between the mobile station MS and the radio network UTRAN and for transmitting control information from the core network CN and the radio network RAN to the mobile stations MS. In addition, the radio resource control protocol RRC is responsible for allocating enough capacity for the radio bearer according to the instructions of the radio resource management system RRM in application-based capacity allocation, for instance.

[0017] A protocol stack as shown in Figure 2b is used in transmitting UMTS packet-switched user data. On the interface Uu between the radio network UTRAN and a mobile station MS, the lower-level data transmission on a physical layer is performed according to a WCDMA or TD-CDMA protocol. A MAC layer above the physical layer transmits data packets between the physical layer and an RLC layer and the RLC layer handles the logical management of the radio links of different radio bearers. The RLC functions comprise for instance segmenting the user data (RLC-SDU) being transmitted into one or more RLC data packets RLC-PDU. Data packets (PDCP-PDU) of a PDCP layer above the RLC, including the header fields of upper layers, can optionally be compressed. After this, the PDCP-PDUs are forwarded to the RLC and they correspond to one RLC-SDU. The user data and the RLC-SDUs are segmented and transmitted in RLC frames, to which address and verification information essential for data transmission is added. The RLC layer also takes care of re-transmission of damaged frames. The serving node 3G-SGSN manages the routing of the data packets coming from

the mobile station MS through the radio network RAN on to the correct gateway node 3G-GGSN. This connection uses the tunneling protocol GTP which encapsulates and tunnels all user data and signaling transmitted through the core network. The GTP protocol runs on top of the IP used by the

5

[0018] One of the tasks of the PDCP layer is to enable a transparent transmission of data packets coming from the higher application-level layers on to the lower link-level layers and vice versa between UMTS terminals and elements of the radio network UTRAN. Thus, it must be possible

10

to modify the PDCP layer in such a manner that it can also transmit data packets of other network-level protocols than the IP protocols (IPv4, IPv6) already supported.

[0019] Another important task of the PDCP layer comprises functions related to improving channel efficiency. These functions are typically

15

based on different optimization methods, such as compression algorithms of data packet header fields. Since today the network-level protocols planned for the UMTS are IP protocols, the compression algorithms used are those standardized by the IETF (Internet Engineering Task Force). It is, however, possible to apply to the PDCP layer any header compression algorithm which

20

is selected according to the network-level protocol used in each case. Some header compression algorithms may require a bi-directional connection between the terminal and the network so that it is possible to take care of various acknowledgments and manage recovery from errors.

25

[0020] The tasks of the PDCP layer also include transmitting data packets PDCP-SDU and the related PDCP sequence numbers to a new radio network sub-system in internal handovers between UMTS radio network sub-systems (SRNS Relocation). Another task is to multiplex several radio bearers to one and the same RLC entity, when necessary.

30

[0021] Figure 3 shows an functional model of the PDCP layer, in which one PDCP entity is defined for each radio bearer. Since in the present systems, an individual PDP context is defined for each radio bearer, one PDCP entity is also defined for each PDP context, and a certain RLC entity is defined for each PDCP entity on the RLC layer. As stated above, the PDCP layer can in principle be operationally implemented in such a manner that

35

several PDP contexts are multiplexed on the PDCP layer, in which case on the

RLC layer below the PDCP layer, one RLC entity receives data packets from several radio bearers at the same time.

[0022] Each PDCP entity can use one or more header compression algorithms or not use any. Several PDCP entities can also use the same algorithm. The radio resource controller RRC negotiates a suitable algorithm for each PDCP entity as well as parameters controlling the algorithm and then advises the selected algorithm and parameters to the PDCP layer through a PDCP-C-SAP point (PDCP Control Service Access Point). The used compression method depends on the network-level protocol type used on the connection, the type being indicated to the radio resource controller when the PDP context is activated.

[0023] Indicating and distinguishing various compression methods on the PDCP layer is done by means of packet identifiers PID attached to the data packets PDU. For the packet identifier PID values of each PDCP entity a table is created, in which different compression algorithms are matched to different data packets and the value of the packet identifier PID is determined as a combination of these. If no compression algorithm is used, the packet identifier PID obtains the value zero. For each compression algorithm and its combinations with different data packet types, PID values are determined sequentially in such a manner that the PID values of each compression algorithm start from $n+1$, wherein n is the last PID value defined for the previous compression algorithm. The order of the compression algorithms is determined in negotiations with the radio resource controller RRC. The PDCP entities at each end of the packet data connection can identify the compression algorithms of transmitted and received data packets on the basis of the PID value table. This information is, however, not stored in the radio resource controller RRC.

[0024] In application-based capacity allocation, in which for instance an application of a mobile station MS transmits a request to the network to establish a radio bearer, a capacity request is transmitted from the mobile station MS as control signaling to the function SM (Session Management) managing core network connections, from which the capacity request is forwarded to the corresponding function SM of the serving node 3G-SGSN. The serving node 3G-SGSN negotiates with the radio resource management system RRM of the radio network controller RNC on whether radio resources according to the capacity request are available. If there are

enough resources, the serving node 3G-SGSN gives the resource allocation task to the radio resource management system RRM whose tasks include the allocation of the limited radio resources as optimally as possible to the different radio bearers. The radio resource management system RRM
 5 determines what kind of radio resource parameters would be optimal for using the application and defines according to the available radio resource capacity the most suitable parameters for the radio bearer. The radio resource management system RRM transmits instructions to the radio resource control protocol RRC which performs the actual radio resource allocation. For
 10 applications which require a one-way connection only, all the available capacity, for instance x kbit/s, is typically allocated for the one direction, typically the downlink direction, and no capacity, i.e. 0 kbit/s, is allocated for the other direction, i.e. the uplink direction. If the terminal then tries or is forced to use a header compression algorithm requiring a bi-directional connection,
 15 data transmission between the network and the terminal will not succeed.

[0025] Now by means of the present invention, this can be avoided in such a manner that the compression method used for the radio link is selected already before establishing the radio bearer and the information is transmitted to the radio resource management system RRM which takes into
 20 account the possibly selected bi-directional compression algorithm in capacity allocation. In this case, the total bandwidth allocated for the radio bearer can be optimized by taking into account both the bandwidth required by the compression method and the benefit derived from the use of the compression algorithm as a smaller amount of data.

[0026] This can be illustrated by means of Figure 4 which shows the signaling of an embodiment for allocating capacity for a bi-directional connection. The PDCP layer of the user equipment UE supports at least one compression algorithm which requires a bi-directional connection. Information on the compression algorithms (400) supported by the user equipment UE is
 30 transmitted, for instance in a UE_capability message known per se, to the radio resource control layer RRC of the radio network controller RNC when the radio bearer is established. The radio network controller RNC decides (402) the compression algorithm to be used on the radio bearer and advises this to the radio resource management system RRM (404). In response to an application-based capacity allocation request, the serving node 3G-SGSN and
 35 the radio resource management system RRM of the radio network controller

RNC negotiate as described above whether there are enough radio resources. If enough resources are available, the RRM allocates capacity to the radio bearer in such a manner that an as optimal bandwidth as possible is defined based on the application, also taking into account, however, possible
 5 limitations set by the compression algorithm. On the basis of these conditions, the RRM decides how the radio resource controller RRC is to be configured and informs the configuration instructions to the RRC (406). On the basis of the configuration, the RRC makes the final capacity allocation (408) to the radio bearer in question, which allocation also ensures that there is enough
 10 capacity allocated for the reverse connection, if necessary.

[0027] The above procedure can be illustrated by means of the following example. The user of a terminal wants to download a video file from the network, and due to the application used to play the video file, the necessary data transmission rate in the downlink direction is 100 kbit/s and in
 15 the uplink direction 0 kbit/s. On the basis of the UE_capability message transmitted by the terminal, the radio resource controller RRC notices that the PDCP entities of both the terminal and the base station support a header compression algorithm according to the Internet standard proposal RFC2507, which requires a bi-directional connection. The RRC selects said compression
 20 algorithm for the radio bearer and advises it to the radio resource management system RRM. Using said compression algorithm requires a data transmission rate of 5 kbit/s, for instance, in the uplink direction. If the proportion of the header field of the entire data transmission is estimated to be 30 kbit/s, for instance, (whereby the proportion of the payload data is 70 kbit/s)
 25 and the proportion of the header field after header compression is 10 kbit/s, for instance, the data transmission rate of the downlink direction can be set at 80 kbit/s. This way, the RRM sets the radio resource controller RRC to be configured in such a manner that 80 kbit/s is allocated for the downlink and 5 kbit/s is allocated for the uplink, whereby the operation of the desired
 30 compression method is ensured and the 20 kbit/s saved in the downlink direction with respect to the application-based capacity allocation can be allocated to another user.

[0028] In the above, the invention has been described in connection with wireless packet-switched data transmission, especially with the radio
 35 resources of the UMTS system. The invention is, however, not limited to wireless data transmission only, but can also be utilized in packet-switched

data transmission by wire, employing application-based data transmission capacity allocation. On connections based on the Internet protocol (IP), such as TCP (Transmission Control Protocol) or UDP (User Datagram Protocol) connections, the possibly used header compression methods are then indicated to the receiving party and data transmission resources are allocated to the terminal connection taking into account the capacity required by the header compression methods in the allocated data transmission resources.

[0029] It is obvious to a person skilled in the art that while technology advances, the basic idea of the invention can be implemented in many different ways. The invention and its embodiments are thus not restricted to the examples described above, but can vary within the scope of the claims.